

NOAA's second-generation global medium-range ensemble reforecast data set

4 Thomas M. Hamill,¹ Gary T. Bates,² Jeffrey S. Whitaker,¹ Donald R. Murray,² Michael
5 Fiorino,³ Thomas J. Galarneau, Jr.,⁴ Yuejian Zhu,⁵ and William Lapenta⁵

⁷ ¹ NOAA Earth System Research Lab, Physical Sciences Division, Boulder, Colorado

*² Cooperative Institute for Research in the Environmental Sciences,
University of Colorado, Boulder*

10 ³*NOAA Earth System Research Lab, Global Systems Division, Boulder, Colorado*

⁴ National Center for Atmospheric Research, Boulder, Colorado

15 Submitted to

16 *Bulletin of the American Meteorological Society*

17 21 September 2012

20 Dr. Thomas M. Hamill

21 NOAA Earth System Research Lab, Physical Sciences Division

22 R/PSD 1, 325 Broadway

23 Boulder, CO 80305

24 e-mail: tom.hamill@noaa.gov

25 phone: (303) 497-3060

26 fax: (303) 497-6449

27
28

ABSTRACT

29 A multi-decadal ensemble reforecast database is now available that is consistent
30 with the operational 2012 NOAA Global Ensemble Forecast System (GEFS). The
31 reforecast data set consists of an 11-member ensemble run once each day from
32 0000 UTC initial conditions. Reforecasts are run to +16 days. As with the
33 operational 2012 GEFS, the reforecast is run at T254L42 resolution (approximately
34 $\frac{1}{2}$ -degree grid spacing, 42 levels) for week +1 forecasts and T190L42
35 (approximately $\frac{3}{4}$ -degree grid spacing) for the week +2 forecasts. Reforecasts were
36 initialized with Climate Forecast System Reanalysis initial conditions, and
37 perturbations were generated using the ensemble transform with rescaling
38 technique. Reforecast data is available from 1985 to current.

39

40 Reforecast data sets were previously demonstrated to be very valuable for detecting
41 and correcting systematic errors in forecasts, especially forecasts of relatively rare
42 events and longer-lead forecasts. What is novel about this reforecast data set
43 relative to the first-generation NOAA reforecast is that: (a) a modern, currently
44 operational version of the forecast model is used (the previous reforecast used a
45 model version from 1998); (b) a much larger set of output data has been saved,
46 including variables relevant for precipitation, hydrologic, wind-energy, solar-
47 energy, severe weather, and tropical cyclone forecasting; and (c) the archived data
48 is at much higher resolution.

49

50 The article describes more about the reforecast configuration and provides a few

51 examples of how this second-generation reforecast data may be used for research
52 and a variety of weather forecast applications.

53

54

55 CAPSULE SUMMARY
56

57 NOAA's second-generation global ensemble reforecast data set has been created and
58 is freely accessible to the weather forecast community.

59

60

“Those who cannot remember the past are condemned to repeat it.”

- George Santayana

65 1. Introduction.

The weather and climate prediction community have made continued, significant improvement in the quality of numerical forecast guidance. This has come as a result of increased resolution, improved physical parameterizations, improved chemistry and aerosol physics, improved estimates of the initial state estimate due to better data assimilation techniques, and improved couplings between the atmosphere with the land surface, cryosphere, and ocean, and more. Nonetheless, judging from the pace of past improvements, medium-range forecast systematic errors will not become negligibly small within the next decade or two. For intermediate-resolution simulations such as those from current-generation global ensemble systems, users of forecast guidance may notice biased surface temperature forecasts, or precipitation forecasts with insufficient detail in mountainous terrain, and perhaps too much drizzle or too little heavy rain. They may notice over- or underestimated cloud cover, or that near-surface winds are characteristically much stronger than forecast. They may notice that hurricanes are too large in size but less intense than observed. Sometimes, however, systematic errors may be less obvious. Does the model forecast of the Madden-Julian Oscillation (MJO; Zhang 2005) propagate too slowly or decay too quickly? Are Arctic cold outbreaks too intense, and do they plunge south too quickly or too slowly? Does the model over-forecast the frequency of tropical cyclogenesis in the Caribbean Sea? Do tropical cyclones tend to recurve too quickly or slowly? Such

86 questions may be hard to answer quantitatively with a month or even a year of
87 model guidance.

88 In such circumstances, reforecasts can be used to great advantage to
89 distinguish between the random and the model errors. Reforecasts are especially
90 helpful for statistically adjusting weather and climate forecasts to observed data,
91 ameliorating the errors and improving objective guidance (Hamill et al. 2006,
92 Hagedorn 2008). Reforecasts, also commonly called hindcasts, are retrospective
93 forecasts for many dates in the past using the same forecast model and same
94 assimilation system used operationally. Reforecasts have been shown to be
95 particularly useful for the calibration of relatively uncommon phenomena such as
96 heavy precipitation (Hamill et al. 2008) and longer-lead weather-climate
97 phenomena (Hamill et al. 2004), where there is small forecast signal and
98 comparatively large noise due to chaos and model error. In both cases, the large
99 sample size afforded by reforecasts is useful for finding a suitably large number of
100 past similar forecast scenarios. With associated observational data, one then can
101 estimate a conditional distribution of the possible observed states given today's
102 numerical guidance, assuming past forecasts have similar errors to current
103 forecasts. Even when no observed data is available for calibration, reforecasts can
104 be useful for determining the climatology of a model. A 20 ms^{-1} surface wind would
105 be exceptionally strong in most locations on earth, but if the forecast model severely
106 over-forecasts wind speeds, such an event may be of less concern. A reforecast can
107 thus be used for estimating the forecast climatology, placing the current forecast in
108 context (Lalaurette 2003ab).

109 The reforecast data set discussed here makes an unprecedentedly large
110 volume of data accessible to users. Over 27 years of daily, 11-member ensemble
111 forecasts were computed using the currently operational NCEP Global Ensemble
112 Forecast System (GEFS). More than 125 TB of forecast output is conveniently
113 available for fast-access download, and the full model data set (~ 1 PB) is archived
114 on tape. This data set is more extensive than contemporary alternatives, such as the
115 5-member, \sim 20-year, weekly reforecasts from the European Centre for Medium
116 Range Weather Forecasts (ECMWF; Hagedorn 2008, Hagedorn et al. 2011), and
117 there is no charge for its use. Daily lagged reforecasts were also generated for the
118 National Centers for Environmental Prediction (NCEP) Climate Forecast System
119 seasonal forecasts (Saha et al. 2010), though these reforecasts were primarily
120 designed to help the calibration process for forecasts from many weeks to many
121 months lead time.

122 We had several rationales for creating this extensive a reforecast data
123 set. This first is that we hope that the greater number of forecast samples from a
124 statistically consistent model will lead to the diagnosis of model errors and
125 development of novel and improved statistical calibration algorithms and
126 algorithms for rare events and for novel applications, algorithms that may be less
127 accurate were they developed with smaller training data sets. An example of this is
128 products for the renewable energy sector, such as extended-range wind and solar
129 energy potential forecasts. A second major reason is to quantify the benefits of this
130 additional training data. Do we really need an exceptionally large training sample
131 size, or might the products be acceptably similar in skill were they developed with a

132 smaller reforecast data set, perhaps with fewer members or skipping days between
133 reforecasts? Reforecasts are computationally expensive in proportion to the number
134 of prior simulations that were computed. Reforecasts from several decades past
135 may have larger errors due to less accurate initial conditions from the
136 comparatively sparse observational data. Should reforecasting become a regular
137 component of National Weather Service's suite of numerical guidance, it will be
138 helpful to determine the optimal configuration to apply to future ensemble forecast
139 systems, the compromise that provides adequate training data to the statistical
140 applications while being as computationally inexpensive as possible. We also expect
141 now that many users may find other new creative applications for this data set. By
142 making this data freely available, we hope it will be used widely.

143 The next section of the article will discuss the contents of the data set and the
144 procedures to follow in order to download this data. Section 3 will demonstrate
145 some statistical characteristics of the raw reforecast data set. Section 4 describes
146 several forecast applications. Section 5 provides conclusions.

147

148 **2. A description of the reforecast data set and how to access it.**

149 The operational configuration of the NCEP GEFS changed as of 12 UTC 14
150 February 2012. The real-time and reforecast model use version 9.0.1 of the GEFS,
151 discussed at <http://www.emc.ncep.noaa.gov/GFS/impl.php>. For more detail on the
152 GEFS, see Hamill et al. (2011a). During the first eight days of the operational GEFS
153 forecast and the reforecast, the model is run at T254L42 resolution, which with a
154 quadratic Gaussian transform grid is an equivalent grid spacing of approximately 40

155 km at 40° latitude, and 42 vertical levels. Starting at day +7.5, the forecasts are
156 integrated at T190L42, or approximately 54 km at 40° latitude, and data is saved at
157 this resolution from days +8 to days +16, the end of the GEFS integration period.

158 Through 20 February 2011, control initial conditions were generated by the
159 Climate Forecast System Reanalysis, or “CFSR” (Saha et al. 2010). This used the
160 Grid-Point Statistical Interpolation (GSI) System of Kleist et al. (2009) at
161 T382L64. From 20 February 2011 through May 2012, initial conditions were taken
162 from the operational GSI analysis, internally computed at T574L64. After 22 May
163 2012, the GSI was upgraded to use a hybrid ensemble Kalman filter-variational
164 analysis system (Hamill et al. 2011).

165 The perturbed initial conditions for both the operational GEFS and the
166 reforecast use the ensemble transform technique with rescaling, or “ETR” (Wei et al.
167 2008). For the operational real-time forecasts, 80 members are cycled for purposes
168 of generating the initial condition perturbations. However, only the leading 20
169 perturbations plus the control initial condition were used to initialize the
170 operational medium-range forecasts. The operational medium-range GEFS
171 forecasts are generated every six hours from 00, 06, 12, and 18 UTC initial
172 conditions. In comparison, the reforecast was generated only once daily, at 00 UTC,
173 and only 10 perturbed forecast members and the one control forecast were
174 generated. However, the six-hourly cycling of ETR perturbations was preserved,
175 though this cycling used only the 10 perturbed members rather than the 80 used in
176 real time. Model uncertainty in the GEFS is estimated with the stochastic tendencies
177 following Hou et al. (2008) for both operations and reforecasts.

178 Here are some details on the reforecast data that is available. About 27.5
179 years (Dec 1984 - Jun 2012) of reforecast data are currently archived. The archive
180 will soon include the 0000 UTC GEFS real-time forecasts for mid 2012 and beyond,
181 which will be available with some delay, perhaps by 0900 UTC, though many fields
182 will be available more quickly via the NOAA/National Operational Model Archive
183 and Distribution System (NOMADS; nomads.ncdc.noaa.gov). 98 different forecast
184 global fields are available at 1-degree resolution, and 28 selected fields are also
185 available at the native resolution (~0.5-degree resolution on a Gaussian grid for the
186 first week's forecasts, and ~0.67-degree resolution for the second week's forecasts).
187 Data is internally archived in GRIB2 format.

188 (<http://www.nco.ncep.noaa.gov/pmb/docs/grib2/>). The 1-degree data was
189 created from the native resolution data via bilinear interpolation using wgrib2
190 software (<http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/>). The listing
191 of the fields that were saved and their resolutions are provided in Tables 1 and
192 2. Reforecast data was saved at 3-hourly intervals from 0 to 72 h, and every 6 h
193 thereafter. The 27+ years of data daily currently archived totals approximately 125
194 TB of internal storage.

195 Reforecast data can be accessed in many different ways. For users who want
196 a few select fields (say, precipitation forecasts) spanning many days, months, or
197 years, we provide a web interface for accessing such data. The URL is
198 <http://esrl.noaa.gov/psd/forecasts/reforecast2/>. The interface allows the user to
199 select particular fields, date ranges, domains, and type of ensemble information
200 (particular members, the mean, or the spread). While data is internally archived in

201 GRIB2 format, the synthesized files produced from a user's web form input are in
202 netCDF format (<http://www.unidata.ucar.edu/software/netcdf/>) . Should a user
203 desire GRIB2 data instead, the raw data can be accessed via anonymous ftp at
204 <ftp://ftp.cdc.noaa.gov/Projects/Reforecast2> or using wgrib2's "fast downloading"
205 capabilities
206 (www.cpc.ncep.noaa.gov/products/wesley/fast_downloading_grib.html). We
207 request that users be conservative with their downloads in order to minimize
208 computations and bandwidth.

209 Some users may desire only selected days of reforecasts but want full model
210 output rather than the limited set of fields and levels available from ESRL. In this
211 case, the user can download this data from the tape archive at the US Department of
212 Energy. The web form for this is at
213 <http://portal.nersc.gov/project/refcst/v2/>. Such full data may be useful for, say,
214 initializing high-resolution regional reforecasts. An example of this will be
215 provided in section 4.

216

217 **3. Characteristics of the raw reforecast data.**

218 The skill of the second-generation global ensemble reforecasts has improved
219 very significantly from those from the first generation. Figure 1 shows a time series
220 of yearly-averaged global 500-hPa geopotential height anomaly correlations (AC)
221 from both systems. For recent years, the day +5 second-generation reforecasts are
222 more accurate than the day +3 first-generation reforecasts. Considering the
223 second-generation reforecast, there is a modest change in average skill of the

224 reforecasts during the 26-year period shown. Yearly average AC increases in the
225 version 2 reforecasts during the period with the change somewhat less than one
226 day. For example, the day+5 forecasts for 2009-2010 appear to be roughly
227 comparable to the day +4 forecasts (not shown) from 1985-1986.

228 Tropical cyclone forecast tracks were calculated using the GFDL tracker
229 algorithm (Gopalakrishnan et al. 2012). Figure 2 shows track statistics binned by
230 half decades. There has been a pronounced improvement in track forecasting
231 during the period of the reforecasts. This is at least in part due to greater changes in
232 the forecast skill of the steering flow in the tropics, due to improvements in the
233 CFSR analyses over time. 500 hPa geopotential height anomaly correlations
234 improved by 1-2 days between 1985-1986 and 2009-2010 (not shown). Such large
235 changes in skill during the reforecast period can make it more difficult to achieve
236 high forecast accuracy with simple statistical post-processing algorithms, for the
237 forecast errors in past cases will not be fully representative of current forecast
238 errors. Some of these differences, however, also might be due to a change in the
239 accuracy of the observed locations; past observed tracks may not be as accurate as
240 more recent observed tracks. Our own internal computations of blended
241 climatology and persistence track forecasts (CLIPER; Neumann 1972) shows that
242 western Pacific CLIPER track errors have also decreased substantially in the past 25
243 years.

244
245

246 **4. Reforecast applications.**

247 We anticipate that many groups will use this reforecast data set to explore,
248 compare, and validate methods for statistically post-processing the model
249 data. Here we consider the usage of the reforecast for post-processing 24-h
250 accumulated precipitation forecasts, both probabilistic and deterministic.

251 Previously, an analog technique was demonstrated with the first-generation
252 reforecasts as one possible method for statistically downscaling and correcting the
253 forecasts, improving their reliability and skill (Hamill et al. 2006, Hamill and
254 Whitaker 2006). Figure 3 shows Brier Skill Scores from the first- and second-
255 generation reforecasts, processed using the rank analog technique described more
256 generally in Hamill and Whitaker (2006). Skill scores were calculated in the
257 conventional manner (Wilks 2006), ignoring the tendency to over-forecast skill by
258 not separating the data into subsets with homogeneous climatological uncertainty
259 (Hamill and Juras 2006). Analogs were based on similarities of past forecasts to the
260 current forecast for the current grid point and others in a ~100-km (7x7 grid point)
261 box around the point of interest. Different numbers of analogs were used,
262 depending on how unusual the precipitation forecast was for the day in question.

263 When the event was rather common, judged relative to the forecast climatology, as
264 many as 200 members were used. When the forecast event was in the extreme tail
265 of the forecast distribution, as few as 30 analogs were selected.

266 North American Regional Reanalysis (NARR) 24-h accumulated precipitation
267 analysis data (Mesinger et al. 2006) was used both for training (cross validated by
268 year) and verification. There are systematic errors with the NARR (Bukovsky and

269 Karoly 2007). Still, currently we know of no other precipitation analysis that has
270 the NARR's complete coverage of the contiguous US over the full period of the
271 reforecasts. We use it here, for better and worse.

272 The post-processed forecasts validated from 1985-2010 show an
273 improvement of slightly greater than +1 day additional lead time at the early
274 forecast leads from the first to the second-generation reforecast, i.e., a 24-48 h
275 version 2 forecast could be made as skillfully as the previous 0-24 h forecast from
276 version 1. At longer leads, the improvement sometimes approaches +2 days
277 additional lead time. The improvement of post-processed forecasts from version 1
278 to version 2 is smaller than the improvement in the raw forecast guidance. This is to
279 be expected; the post-processing is correcting more systematic error in version 1
280 than in version 2. Post-processed guidance from both versions are highly reliable,
281 though forecasts from version 2 tend to issue high and low probabilities more
282 frequently, i.e., they are more "sharp" (not shown). Forecast skill probably is over-
283 estimated somewhat for the samples early on in the reforecast period (say, the
284 1980's), for the cross-validated training procedure used analogs from future
285 forecasts that were more accurate.

286 Deterministic forecasts can also be improved with the statistical post-
287 processing. A slightly different approach was used to generate the deterministic
288 forecast from the analogs. First, rather than using the observed on days with similar
289 forecasts, the difference between observed minus forecast on the days with the
290 closest analog forecasts was used to "dress" the current forecast; this provided
291 somewhat higher precipitation amounts when anomalously large events were

292 forecast. The mean of this dressed set of analog forecasts was then computed. As
293 with deterministic forecasts generated from an ensemble-mean forecast, the analog
294 mean forecast tends to over-forecast the light precipitation and under-forecast
295 heavy precipitation. Hence, following Ebert's probability-matched mean approach
296 (<http://www.cawcr.gov.au/staff/eee/etrap/probmatch.html>) the ensemble mean of
297 the analogs was adjusted before it was used as a deterministic forecast. Specifically,
298 for all the forecasts for a given month of the year, the cumulative distribution
299 function (CDF) of these analog ensemble-mean forecasts was computed (cross-
300 validated) using the current month and the surrounding two months, as well as the
301 CDF of the NARR data set. The quantile associated with the current analog mean
302 forecast relative to the forecast climatology was noted, and the final deterministic
303 forecast was the precipitation amount associated with the corresponding analyzed
304 quantile. Figure 4 shows that the analog post-processed deterministic forecast skill
305 also provides an improvement relative to either the GEFS control or ensemble mean,
306 particularly at the light precipitation amounts, where apparently there was a drizzle
307 over-forecast bias. The ensemble mean from the raw ensemble shows a
308 characteristic under-forecast bias, while the control forecast has a slight over-
309 forecasting bias.

310 These calibration approaches are relatively simple; they are univariate,
311 based only on the forecast precipitation amount. Also, they don't factor in changes
312 in skill of the forecasts during the training period. We hope and expect that other
313 groups will explore methods that may extract further value from the extensive
314 reforecast data set, using new techniques and additional predictors.

315 Suppose a long time series of observations is not available to accompany the
316 time series of reforecasts. How can one leverage the reforecasts to provide value-
317 added guidance? Reanalyses might be used for the calibration, but analyses may be
318 contaminated somewhat by model forecast bias. Should the user desire guidance
319 for a point location, the reanalysis cannot provide this, only the grid-box averaged
320 analyzed state. In such cases, perhaps usage of diagnostics like the Extreme
321 Forecast Index (EFI; LaLaurette 2003ab) may be of use. The EFI quantifies how
322 unusual the current ensemble guidance is relative to the climatology of past forecast
323 guidance. Ideally, even when the ensemble guidance is biased in some fashion, it
324 can still provide some advanced warning of potential extreme events. For such
325 events, today's ensemble guidance should be ranked in the extreme quantiles of the
326 distribution defined by the past forecasts.

327 Figure 5 considers the problem of extended range wind-energy forecasts,
328 specifically a +5 to +10 day forecast of 80-m above ground level wind speeds, a
329 common height of the hubs of wind turbines. Suppose a wind farm operator in
330 North Dakota does not have a multi-decadal time series of wind observations at hub
331 height, but they wish to extract some information from a reforecast that may
332 indicate when it would be relatively inexpensive to shut down a turbine for
333 maintenance. Figure 5(a) shows the ensemble mean forecast wind speed for a
334 particular case day in early 2010. The winds appear relatively light on average in
335 this location, but they might be biased. However, the availability of the reforecasts
336 allows that wind speed forecast to be placed in context. Figure 5(b) shows the
337 quantile of the ensemble-mean forecast wind speed relative to its climatology for

338 that month, a calculation similar in spirit to the EFI. The wind speed forecasts are
339 indeed unusually light in this location relative to their forecast climatology, which
340 ended up being consistent with analyzed conditions (Figs. 5 c,d).

341 Let's turn our attention from post-processing to other potential applications
342 of the reforecasts. One possible application is to use the global reforecast ensemble
343 data as initial and lateral boundary conditions for a high-resolution regional
344 reforecast ensemble. The ability to perform high-resolution regional reforecasts
345 may be of interest to many, perhaps to examine the ability of a higher-resolution
346 regional model to provide value-added guidance for high-impact weather events. As
347 discussed in section 2, the full model output for the global reforecast ensemble is
348 available on the U.S. Department of Energy website. An illustrative example of a
349 regional reforecast ensemble is shown in Fig. 6. Here, an 11-member ensemble 72-
350 hour forecast initialized at 0000 UTC 22 September 2005 for TC Rita was generated
351 using version 3.3 of the Advanced Hurricane Weather Research and Forecasting
352 (ARW) model, with 36 vertical levels up to 20 hPa (Skamarock et al. 2008). Details
353 of the modification of ARW for hurricane applications are described in Davis et al.
354 (2008). This implementation of ARW was run over a fixed 36-km domain that
355 covers the entire North Atlantic basin, North America, and the extreme eastern
356 North Pacific (see Fig. 2 and Table 1 in Galarneau and Davis (2012)). Two-way
357 moving nests of 12- and 4-km are located within the 36-km domain, and the
358 movement of these nests is determined by the TC's motion during the previous 6 h.
359 Specifics on the AHW configuration are as follows: WRF single-moment 6-class
360 microphysics (Hong et al. 2004), modified Tiedtke convective parameterization

361 (Zhang et al. 2011) on the 36- and 12-km domains (no parameterization on the 4-
362 km domain), Yonsei University boundary layer scheme (Hong et al. 2006), Goddard
363 shortwave scheme (Chou and Suarez 1994), Rapid Radiative Transfer Model
364 (Mlawer et al. 1997), and Noah land surface model (Ek et al. 2003).

365 The global reforecast ensemble shows a range of possible model trajectories,
366 including significant impact on Houston, Texas (Fig. 6a). The track forecast from the
367 global reforecast ensemble was consistent with the official National Hurricane
368 Center track forecast for Rita three days prior to landfall (not shown), which
369 resulted in an evacuation order for the Houston area. The track forecast had a
370 significant left-of-track error, as the observed storm made landfall farther northeast,
371 near the Texas-Louisiana border. The intensity forecast was consistently
372 underestimated in the global reforecast ensemble (Fig. 6a inset), a common
373 characteristic with global data assimilation and forecast systems with grid spacing
374 of many tens of km. The ARW regional reforecast ensemble also had a left-of-track
375 forecast error, although the ensemble track envelope expanded slightly farther
376 northeast along the Gulf coast (Fig. 6b). That the left-of-track error appears in the
377 ARW reforecast ensemble in addition to the global model suggests that track errors
378 were driven by errors in the TC steering flow. This is modulated by large-scale
379 features such as the subtropical ridge over the southeast U.S. and an eastward-
380 moving mid-latitude trough over the central Great Plains (not shown). The ARW
381 reforecast ensemble inherited the initial under-estimate of intensity seen in the
382 global reforecast (Fig. 6a inset), but was able to intensify the storm to a major
383 hurricane by 48-h, just prior to landfall (Fig 6a and 6b insets).

384 Another potential application for reforecasts is to understand the ability of
385 the model to predict uncommon phenomena, or even the relationships between
386 several uncommon phenomena. As an example, let's say that we wanted to
387 understand whether atmospheric blocking statistics (Tibaldi and Molteni 1990) can
388 be correctly forecast given a recently strong or weak MJO. To make the problem
389 more statistically challenging, let's further suppose we are interested in the blocking
390 forecasts related to a certain phase of the MJO, where it is most pronounced in the
391 Indian Ocean, and at a certain time of the year, here December - January - February
392 (DJF). In such a situation, a year or two of past recent forecasts will not provide
393 enough samples.

394 Using the first two empirical orthogonal functions of MJO variability
395 (Wheeler and Hendon 2004), commonly known as RMM_1 (Real-time Multivariate
396 MJO) and RMM_2 , a strong MJO, should it exist, would be classified as being in the
397 Indian Ocean roughly if $RMM_1 \cong 0$ and $RMM_2 << 0$. Accordingly, for the angle θ
398 defined by the arctangent of RMM_1 and RMM_2 , we define the Indian Ocean "strong
399 MJO" as occurring if $-(\pi/2 + \pi/8) \leq \theta \leq -\pi/2 + \pi/8$, and if the amplitude
400 $(RMM_1^2 + RMM_2^2)^{1/2}$ is in the upper quartile of the climatology of analyzed
401 amplitudes for this phase and for DJF. Figure 7(a) shows the CFSR analyzed
402 unconditional Dec-Jan-Feb 1985-2010 blocking statistics and the blocking statistics
403 under a strong Indian Ocean MJO six days prior to the analysis. The lagged
404 observed blocking frequency from the Pacific to the Atlantic Ocean is apparently
405 strongly suppressed with strong MJOs relative to the climatology. Composites (not
406 shown) indicate that there are generally negative 500 hPa height anomalies in the

407 climatological ridges and positive anomalies in the troughs, resulting in generally
408 more zonal flow and less blocking. Fig. 7b shows the blocking frequency in the +6
409 day control member reforecasts (using analyzed RMM_1 and RMM_2 , i.e., a -6 day lag
410 so that analyzed data is used to define the MJO indices). There is a similar
411 depression of the forecast blocking frequency under a strong MJO; the forecast
412 model does well at replicating the climatology of blocking and its relationship to this
413 phase of the MJO. This simple illustration shows how the reforecast data set offers
414 a unique opportunity to potentially diagnose and examine model systematic
415 forecast characteristics related to infrequent or low-frequency phenomena.

416

417 **5. Conclusions.**

418 For the foreseeable future, weather and climate prediction model guidance
419 will be contaminated by at least some systematic errors. Since most end users want
420 reliable and accurate guidance, some statistical post-processing may be
421 helpful. Sometimes, such as for rare events and longer-lead forecasts, a long
422 training data set of “reforecasts” can be especially helpful in detecting the useful
423 signal within the bath of chaotic noise and model error.

424 This article described one such data set, a second-generation reforecast that
425 is consistent with the NCEP Global Ensemble Forecast System as it was configured
426 in 2012. We showed a variety of uses of this reforecast data set, such as the
427 statistical post-processing of precipitation forecasts, the initialization of regional
428 reforecasts, and the diagnosis of the forecastability of uncommon phenomena.

429 This data set was generated from a large high-performance computing grant

430 by the U.S. Department of Energy to explore the potential for improving longer-lead
431 weather forecasts related to renewable energy; it was not created on NOAA
432 computers. Currently, NCEP has not allocated any of its high-performance
433 computing to the generation of reforecasts specific to weather time scales. While
434 we intend to keep running this version of the GEFS for the foreseeable future, even
435 after NCEP upgrades its GEFS, the regrettable truth is that soon enough the GEFS
436 will change and the reforecast will be inconsistent with the operational version of
437 the model. ECMWF embraced some years ago the approach of computing a more
438 limited set reforecasts on their operational computer using whatever model version
439 is currently operational. In this way, their reforecast data set is continually relevant
440 to today's model guidance. As NOAA determines the amount of high-performance
441 computing it needs in the coming years and decades, we hope that the computers
442 will be sized so that NOAA too can generate reforecasts on the fly, save the data, and
443 make it readily available to the weather enterprise. This data set, we expect, will
444 help NOAA decide on a realistic configuration for such reforecasts.

445

446 **Acknowledgments:**

447 The U.S. Department of Energy provided the high-performance computing to
448 produce this data set, under its Advanced Scientific Computing Research (ASCR)
449 Leadership Computing Challenge. (ALCC). We are grateful to the DOE and its
450 professional support staff for their help. The mass storage array within ESRL was
451 partially supported by NOAA THORPEX funds distributed by NOAA's Office of
452 Weather and Air Quality (OWAQ). We had tremendous help from the IT staff in the

453 Physical Sciences Division at ESRL, in particular Nick Wilde, Alex McColl, Barry
454 McInnes, Chris Kreutzer, and Eric Estes were all helpful in configuring the storage
455 array and helping us get to voluminous reforecast data to and from it. The ARW
456 reforecast ensemble was generated using the Bluefire supercomputer at the
457 National Center for Atmospheric Research (NCAR). NCAR is sponsored by the
458 National Science Foundation.

459

460

461

462

463

464

465

466

467

468

469 **References**

- 470 Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the
471 North American Regional Reanalysis. *J. Hydrometeor.*, **8**, 837-846.
- 472 Chou, M. -D., and M. J. Suarez, 1994: An efficient thermal infrared radiation
473 parameterization for use in general circulation models. Tech. Rep., NASA
474 Tech. Memo. 104606, 3, 85 pp.
- 475 Davis, C. A., and Coauthors, 2008: Prediction of landfalling hurricanes with the
476 Advanced Hurricane WRF model. *Mon. Wea. Rev.*, **136**, 1990–2005.
- 477 Ek, M. B., and Coauthors, 2003: Implementation of Noah land surface model
478 advances in the National Centers for Environmental Prediction operational
479 mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851.
- 480 Galarneau, T. J., Jr., and C. A. Davis, 2012: Diagnosing forecast errors in tropical
481 cyclone motion. *Mon. Wea. Rev.*, in press.
- 482 Gopalakrishnan, S. and others, 2012: Hurricane Weather Research and Forecasting
483 (HWRF) Model: 2012 Scientific Documentation, Development Testbed
484 Center, Boulder CO
485 (http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HW
486 RFScientificDocumentation_v3.4a.pdf), see pages 71-91.
- 487 Hagedorn, R., 2008: Using the ECMWF reforecast data set to calibrate EPS
488 reforecasts. *ECMWF Newsletter*, **117**, 8-13.
- 489 Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2011:
490 Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF

- 491 ensemble forecasts. *Quart. Journ. of the Royal Meteor. Soc.*, in press. Available at
492 <http://onlinelibrary.wiley.com/doi/10.1002/qj.1895>
- 493 Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: improving
494 medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**,
495 1434-1447.
- 496 Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: an important dataset
497 for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33-46.
- 498 Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation
499 forecasts based on reforecast analogs: theory and application. *Mon. Wea. Rev.*,
500 **134**, 3209-3229.
- 501 Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the
502 varying climatology? *Quart. J. Royal Meteor. Soc.*, **132**, 2905-2923.
- 503 Hamill, T. M., R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration
504 using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea.
505 Rev.*, **136**, 2620-2632.
- 506 Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011a: Global Ensemble
507 Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman
508 Filter. *Mon. Wea. Rev.*, **139**, 668-688.
- 509 Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b:
510 Predictions of 2010's tropical cyclones using the GFS and ensemble-based data
511 assimilation methods. *Mon. Wea. Rev.*, **139**, 3243-3247.

- 512 Hong, S. -Y., J. Dudhia, and S. -H. Chen, 2004: A revised approach to ice microphysical
513 processes for the bulk parameterization of clouds and precipitation. *Mon. Wea.
514 Rev.*, **132**, 103–120.
- 515 Hong, S. -Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an
516 Explicit Treatment of Entrainment Processes. *Mon. Wea. Rev.*, **134**, 2318–2341.
- 517 Hou, D., Z. Toth, Y. Zhu, and W.. Yang, 2008: Impact of a stochastic perturbation
518 scheme on NCEP Global Ensemble Forecast System. *Proceedings, 19th AMS
519 Conference on Probability and Statistics*. New Orleans, LA, 20-24 Jan. 2008.
- 520 Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009:
521 Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea.
522 Forecasting*, **24**, 1691-1705.
- 523 Lalaurette, F., 2003a: Early detection of abnormal weather conditions using a
524 probabilistic extreme forecast index. *Quart. Journ. Royal Meteor. Soc.*, **129**,
525 3037-3057.
- 526 ——, 2003b: Two proposals to enhance the EFI response near the tails of the
527 climate distribution. 8.
- 528 Mesinger, F., and co-authors, 2006: North American Regional Reanalysis. *Bull. Amer.
529 Meteor. Soc.*, **87**, 343-360.
- 530 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997:
531 Radiative transfer for inhomogeneous atmosphere: RRTM, a validated
532 correlated-k model for the long-wave. *J. Geophys. Res.*, **102**, 16663–16682.
- 533 Neumann, C. J., 1972: An alternate to the HURRAN tropical cyclone forecast system.
534 *NOAA Tech. Memo. NWS SR-62*, 22pp. Available from the National Technical

- 535 Information Service, U.S. Department of Commerce, 5285 Port Royal Rd.,
536 Springfield, VA 22151.
- 537 Saha, S., and co-authors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull.
538 Amer. Meteor. Soc.*, **91**, 1015-1057.
- 539 Skamarock, W. C., and Coauthors, 2008: *A Description of the Advanced Research WRF
540 Version 3*. NCAR Tech. Note NCAR/TN-475+STR, 125 pp.
- 541 Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus
542 A*, **42**, 343-365.
- 543 Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the
544 ensemble transform (ET) technique in the NCEP global operational forecast
545 system. *Tellus A*, **60**, 62-79.
- 546 Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO
547 Index: development of an index for monitoring and prediction. *Mon. Wea. Rev.,*
548 **132**, 1917-1932.
- 549 Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences* (2nd
550 Ed.). Academic Press, 627. pp.
- 551 Zhang, C., 2005: Madden-Julian Oscillation. *Rev. Geophys.*, **43**, RG2003.
- 552 Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary
553 layer clouds over the Southeast Pacific in ARW-WRF using a modified Tiedtke
554 cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513.
- 555
556
557

558 FIGURE CAPTIONS

559

560 **Figure 1:** Running mean (an average over the previous 365 days) of the 500-hPa
561 geopotential height anomaly correlation (AC) from the deterministic control
562 reforecasts. The filled areas denote anomaly correlation from the first-generation
563 GFS reforecast described in Hamill et al. (2006); the bounding lower line denotes
564 the Southern Hemisphere AC, the bounding upper line the Northern Hemisphere
565 AC. Blue indicates day +3 forecasts, pink indicates day+5 forecasts, green indicates
566 day+7 forecasts. The second-generation reforecasts are shown without filled areas;
567 thicker lines denote Northern Hemisphere AC, thinner lines the Southern
568 Hemisphere AC.

569 **Figure 2:** Global tropical cyclone track error (solid lines) and spread (dashed) over
570 ~5 year periods during the reforecast. Statistics were accumulated only for 1 June
571 to 30 November of each year and included data from all basins.

572 **Figure 3:** Brier Skill Scores (*BSS*) of 24-h accumulated precipitation forecasts from
573 1985-2010 over the CONUS, post-processed using the rank analog technique. (a)
574 *BSS* for the > 2.5 mm 24 h⁻¹ event. (b) *BSS* for the > 25 mm 24 h⁻¹ event. Scores are
575 plotted as a function of month of the year and for different forecast lead times from
576 1 to 6 days. Solid lines indicate the scores for the second-generation reforecast (V2),
577 dashed lines for the first-generation reforecast (V1). Black, green, red, blue, purple,
578 and orange lines indicate the respective skills for days +1 to +6.

579 **Figure 4:** Equitable threat scores (ETS) and biases (BIA) for ensemble-mean
580 forecasts, control forecasts and deterministic forecasts generated from post-
581 processed analog ensemble-mean forecasts. Panels (a), (b), and (c) provide ETS for

582 the $> 0.5 \text{ mm } 24\text{h}^{-1}$ event, the $> 5 \text{ mm } 24\text{h}^{-1}$ event, and $> 50 \text{ mm } 24\text{h}^{-1}$ event,
583 respectively. Panels (d), (e), and (f) provide BIA for these respective events.

584 **Figure 5:** (a) +5 to +10 day forecast of ensemble-mean 80-m AGL wind speeds,
585 initialized at 00 UTC on 1 January 2010 for the period 00 UTC 6 January to 11
586 January 2010. (b) Quantile for this ensemble mean forecast relative to the
587 cumulative distribution of past ensemble mean forecasts for the month of
588 January. (c) as in (a), but for CFSR analyzed conditions, and (d) as in (b) but for
589 CFSR analyzed.

590 **Figure 6:** 72-h track forecast for hurricane Rita initialized at 0000 UTC 22
591 September 2005 from the (a) global GFS ensemble reforecast and (b) regional ARW
592 ensemble forecast. The individual ensemble member tracks are shown in gray
593 (control run in green) with red dots marking every 24 hours. The observed track is
594 shown in black with black dots marking every day at 0000 UTC. The inset in (a)
595 shows the intensity forecast for Rita from the global GFS ensemble (gray) and ARW
596 (red). The observed intensity is shown by the blue dashed contour. The black line
597 represents the ensemble mean and the shading encompasses intensity values within
598 the 5% and 95% percentiles. The inset in (b) shows the 48-h forecast composite
599 reflectivity (shaded according to the color bar in dBZ) from the 4-km domain of the
600 control member of the ARW ensemble.

601 **Figure 7:** (a) Observed, and (b) +6 day forecast blocking frequency as a function of
602 latitude for December-January-February 1985-2010 (green lines) and for the subset
603 of cases with an Indian Ocean strong MJO as defined in the text. The MJO data was
604 defined 6 days prior to the analysis or the forecast. Grey area denotes differences

605 that are between the 5th and 95th percentile confidence intervals as determined
606 from a block bootstrap algorithm.

607

608

609

610 **Table 1:** Reforecast variables available for selected mandatory and other vertical
611 levels. F indicates geopotential height, and an X indicates that this variable is
612 available from the reforecast data set at 1-degree resolution; a Y indicates that the
613 variable is available at the native ~0.5 degree resolution. AGL indicates “above
614 ground level.” Hybrid sigma-pressure vertical levels (a very close approximation to
615 sigma levels near the ground) are called “hyb.”

616

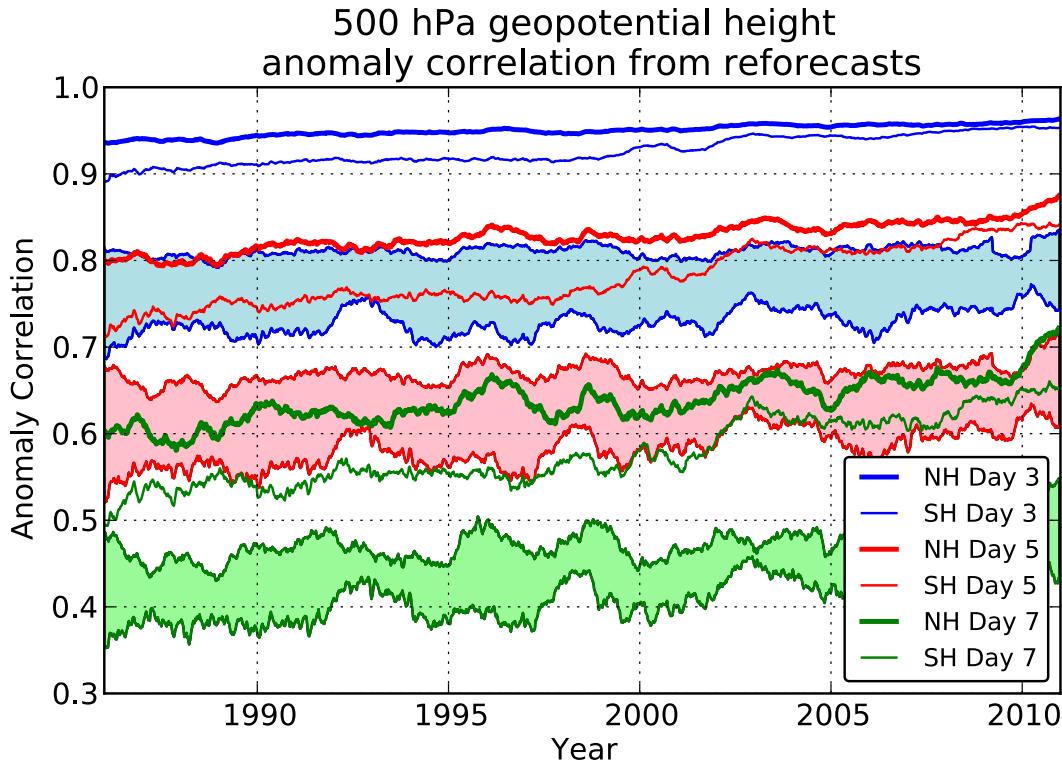
Vertical Level	U	V	T	F	q	Wind Power
10 hPa	X	X	X	X		
50 hPa	X	X	X	X		
100 hPa	X	X	X	X		
200 hPa	X	X	X	X		
250 hPa	X	X	X	X		
300 hPa	X	X	X	X	X	
500 hPa	X	X	X	X	X	
700 hPa	X	X	X	X	X	
850 hPa	X	X	X	X	X	
925 hPa	X	X	X	X	X	
1000 hPa	X	X	X	X	X	
hyb 0.996	X	X		X		
hyb 0.987	X	X		X		
hyb 0.977	X	X		X		
hyb 0.965	X	X		X		
80m AGL	X,Y	X,Y				X,Y

617 **Table 2:** Single-level reforecast variables archived (and their units). Where an [Y]
 618 is displayed, this indicates that this variable is available at the native \sim 0.5-degree
 619 resolution as well as the 1-degree resolution.
 620

Variable (units)
Mean sea-level pressure (Pa) [Y]
Skin temperature (K) [Y]
Soil temperature, 0.0 to 0.1 m depth (K) [Y]
Volumetric soil moisture content 0.0 to 0.1 m depth (fraction between wilting and saturation) [Y]
Water equivalent of accumulated snow depth (kg m^{-2} , i.e., mm) [Y]
2-meter temperature (K) [Y]
2-meter specific humidity (kg kg^{-1} dry air) [Y]
Maximum temperature (K) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]
Minimum temperature (K) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]
10-m u wind component (ms^{-1}) [Y]
10-m v wind component (ms^{-1}) [Y]
Total precipitation (kg m^{-2} , i.e., mm) in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC) [Y]
Water runoff (kg m^{-2} , i.e., mm) [Y]
Average surface latent heat net flux (W m^{-2}) [Y]
Average sensible heat net flux (W m^{-2}) [Y]
Average ground heat net flux (W m^{-2}) [Y]
Convective available potential energy (J kg^{-1}) [Y]

Convective inhibition (J kg^{-1}) [Y]
Precipitable water (kg m^{-2} , i.e., mm) [Y]
Total-column integrated condensate (kg m^{-2} , i.e., mm) [Y]
Total cloud cover (%)
Downward short-wave radiation flux at the surface (W m^{-2}) [Y]
Downward long-wave radiation flux at the surface (W m^{-2}) [Y]
Upward short-wave radiation flux at the surface (W m^{-2}) [Y]
Upward long-wave radiation flux at the surface (W m^{-2}) [Y]
Upward long-wave radiation flux at the top of the atmosphere (W m^{-2}) [Y]
Potential vorticity on the 320K isentropic surface ($\times 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$)
U component on 2 PVU ($1 \text{ PVU} = 1 \times 10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$) isentropic surface (ms^{-1})
V component on 2 PVU isentropic surface (ms^{-1})
Temperature on 2 PVU isentropic surface (K)
Pressure on 2 PVU isentropic surface (Pa)
80-m u wind component (ms^{-1}) [Y]
80-m v wind component (ms^{-1}) [Y]
Vertical velocity at 850 hPa (Pa s^{-1})
Water runoff (kg m^{-2} , i.e., mm)
Wind mixing energy at 80 m (J) [Y]

621
622
623
624
625
626
627



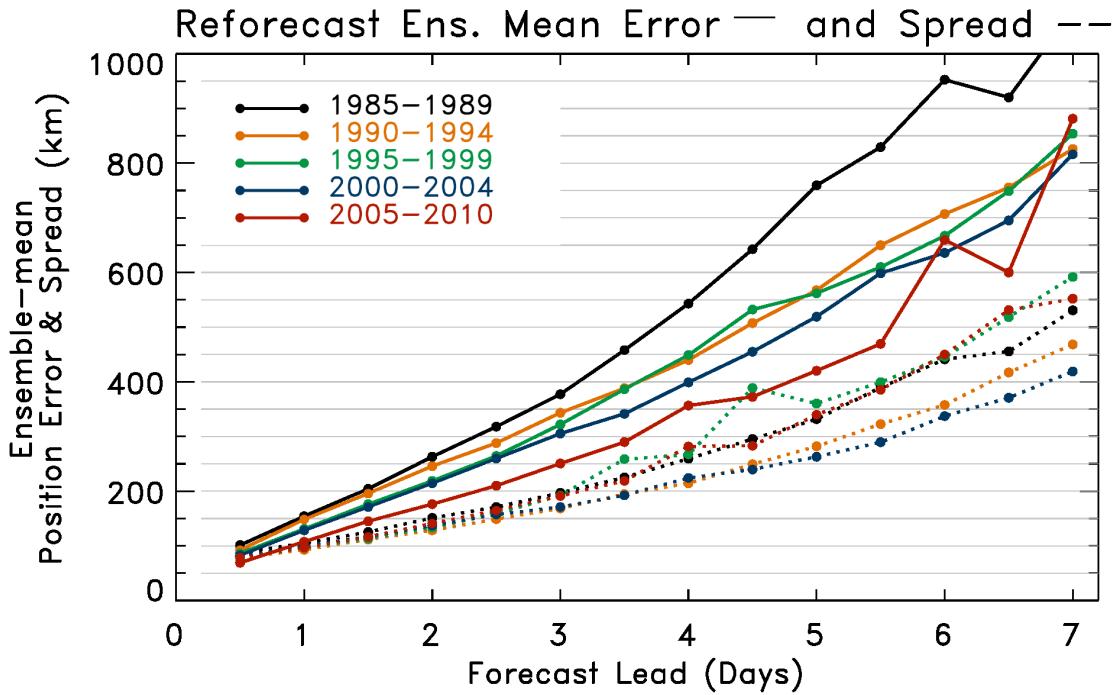
628

629 **Figure 1:** Running mean (an average over the previous 365 days) of the 500-hPa
 630 geopotential height anomaly correlation (AC) from the deterministic control
 631 reforecasts. The filled areas denote anomaly correlation from the first-generation
 632 GFS reforecast described in Hamill et al. (2006); the bounding lower line denotes
 633 the Southern Hemisphere AC, the bounding upper line the Northern Hemisphere
 634 AC. Blue indicates day+3 forecasts, pink indicates day+5 forecasts, green indicates
 635 day +7 forecasts. The second-generation reforecasts are shown without filled areas;
 636 thicker lines denote Northern Hemisphere AC, thinner lines the Southern
 637 Hemisphere AC.

638

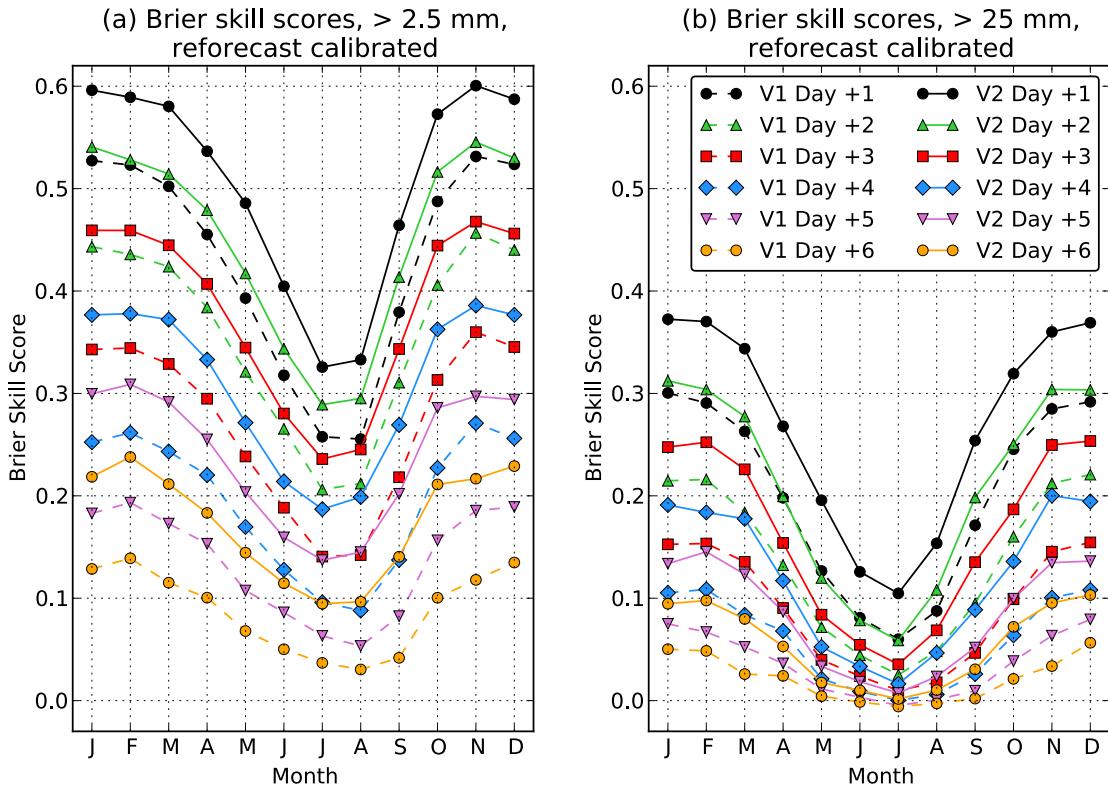
639

640



641
642
643 **Figure 2:** Global tropical cyclone track error (solid lines) and spread (dashed) over
644 ~5 year periods during the reforecast. Statistics were accumulated only for 1 June
645 to 30 November of each year and included data from all basins.

646



647

648 **Figure 3:** Brier Skill Scores (*BSS*) of 24-h accumulated precipitation forecasts from
 649 1985-2010 over the CONUS, post-processed using the rank analog technique. (a)
 650 *BSS* for the > 2.5 mm 24 h⁻¹ event. (b) *BSS* for the > 25 mm 24 h⁻¹ event. Scores are
 651 plotted as a function of month of the year and for different forecast lead times from
 652 1 to 6 days. Solid lines indicate the scores for the second-generation reforecast (V2),
 653 dashed lines for the first-generation reforecast (V1). Black, green, red, blue, purple,
 654 and orange lines indicate the respective skills for days +1 to +6.

655

656

657

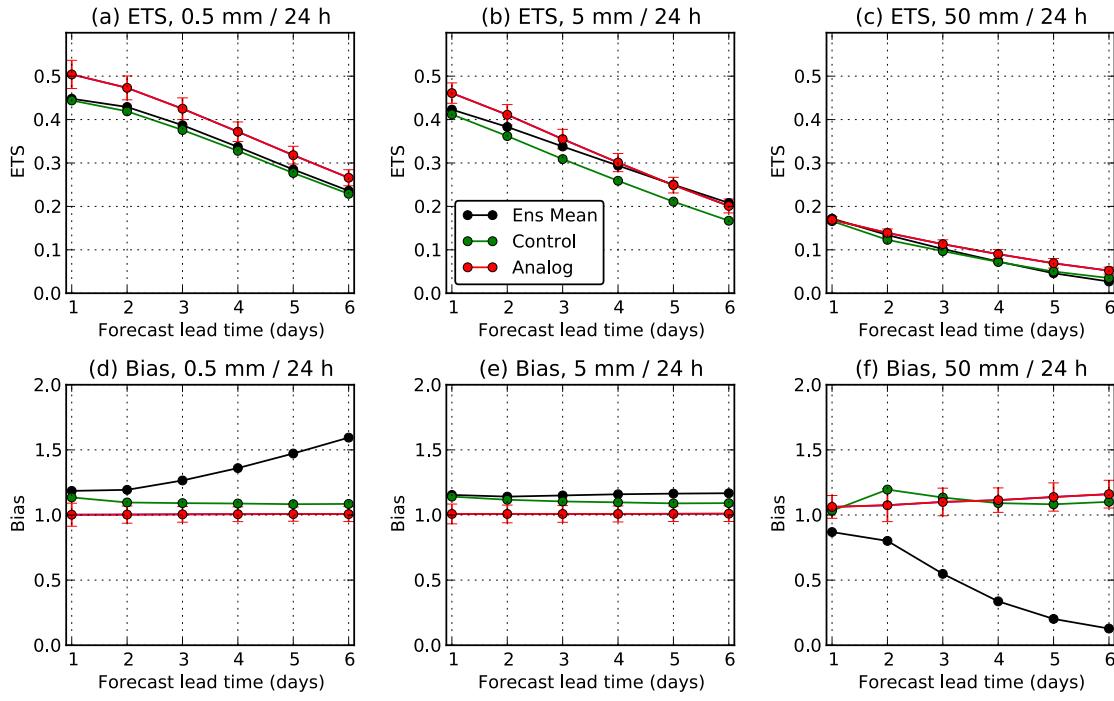
658

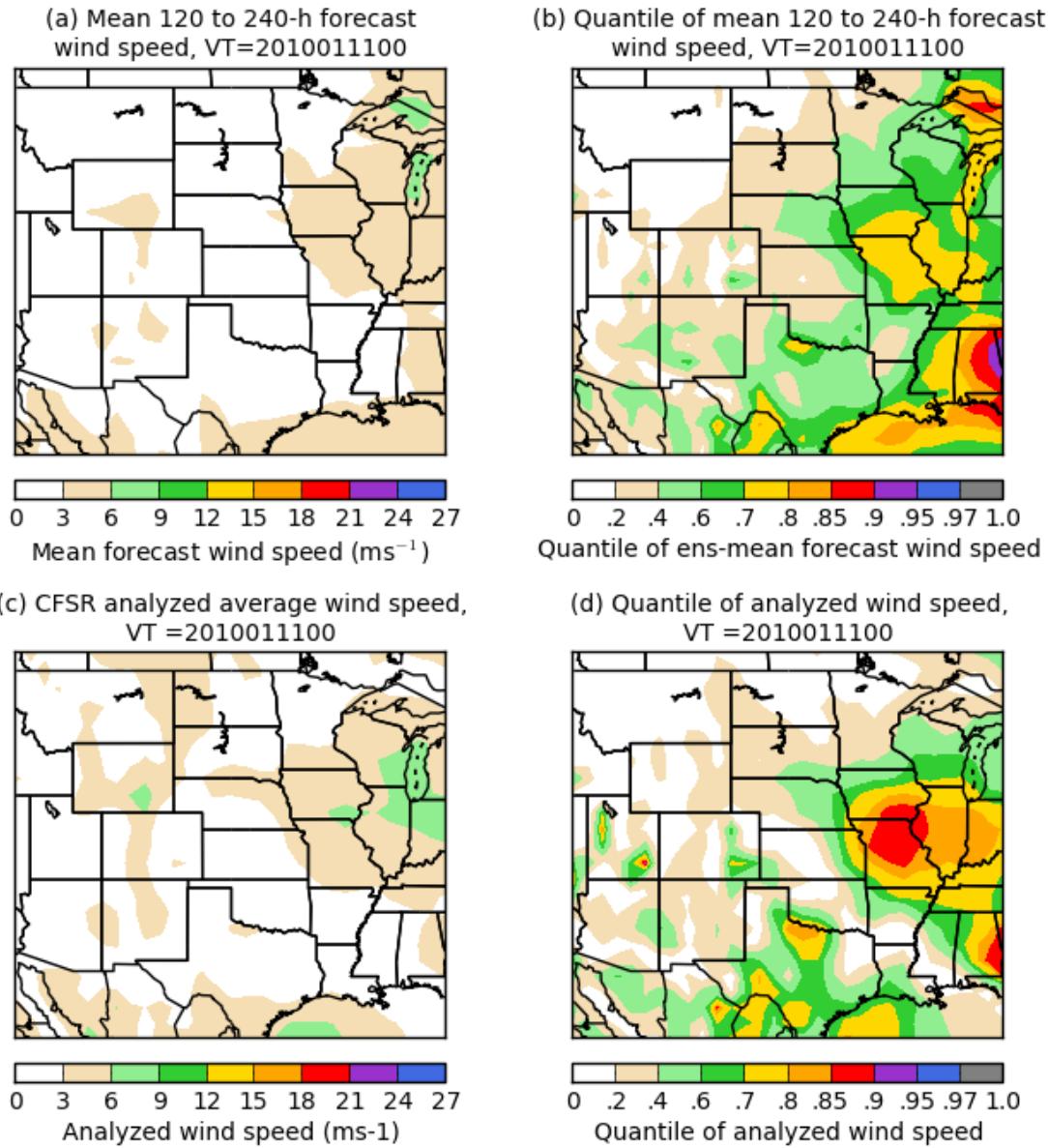
659

660 **Figure 4:** Equitable threat scores (ETS) and biases (BIA) for ensemble-mean
 661 forecasts, control forecasts and deterministic forecasts generated from post-
 662 processed analog ensemble-mean forecasts. Panels (a), (b), and (c) provide ETS for
 663 the $> 0.5 \text{ mm } 24\text{h}^{-1}$ event, the $> 5 \text{ mm } 24\text{h}^{-1}$ event, and $> 50 \text{ mm } 24\text{h}^{-1}$ event,
 664 respectively. Panels (d), (e), and (f) provide BIA for these respective events.

665

666

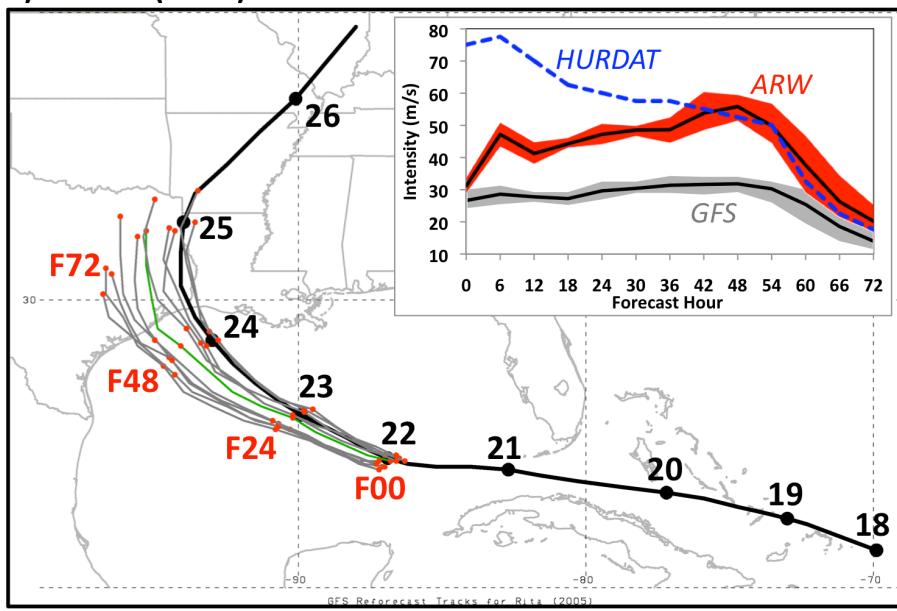




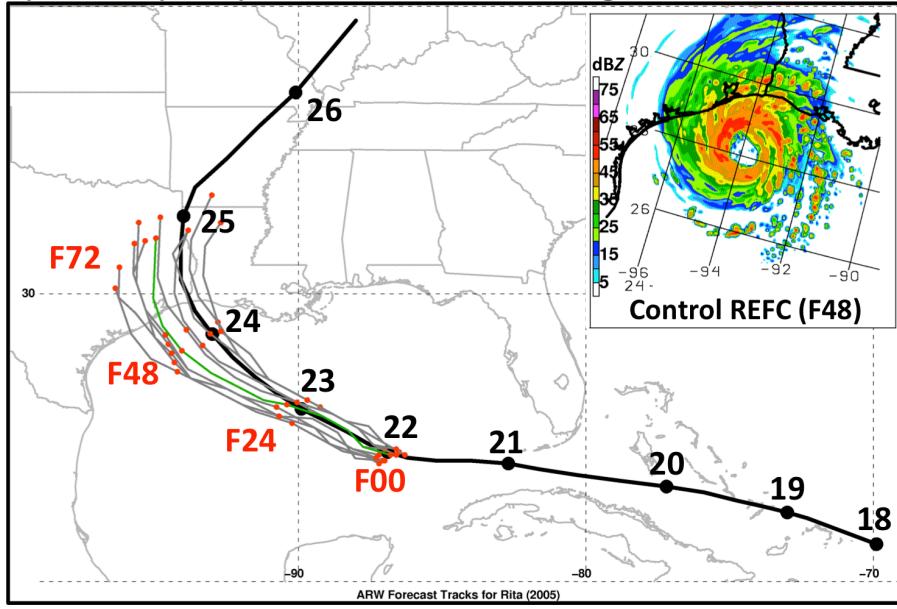
667

668 **Figure 5:** (a) +5 to +10 day forecast of ensemble-mean 80-m AGL wind speeds,
 669 initialized at 00 UTC on 1 January 2010 for the period 00 UTC 6 January to 11
 670 January 2010. (b) Quantile for this ensemble mean forecast relative to the
 671 cumulative distribution of past ensemble mean forecasts for the month of
 672 January. (c) as in (a), but for CFSR analyzed conditions, and (d) as in (b) but for
 673 CFSR analyzed.

a) TC Rita (2005) 72-h GFS Ensemble Reforecast



b) TC Rita (2005) 72-h ARW Ensemble Regional Reforecast



674

675

676 **Figure 6:** 72-h track forecast for hurricane Rita initialized at 0000 UTC 22

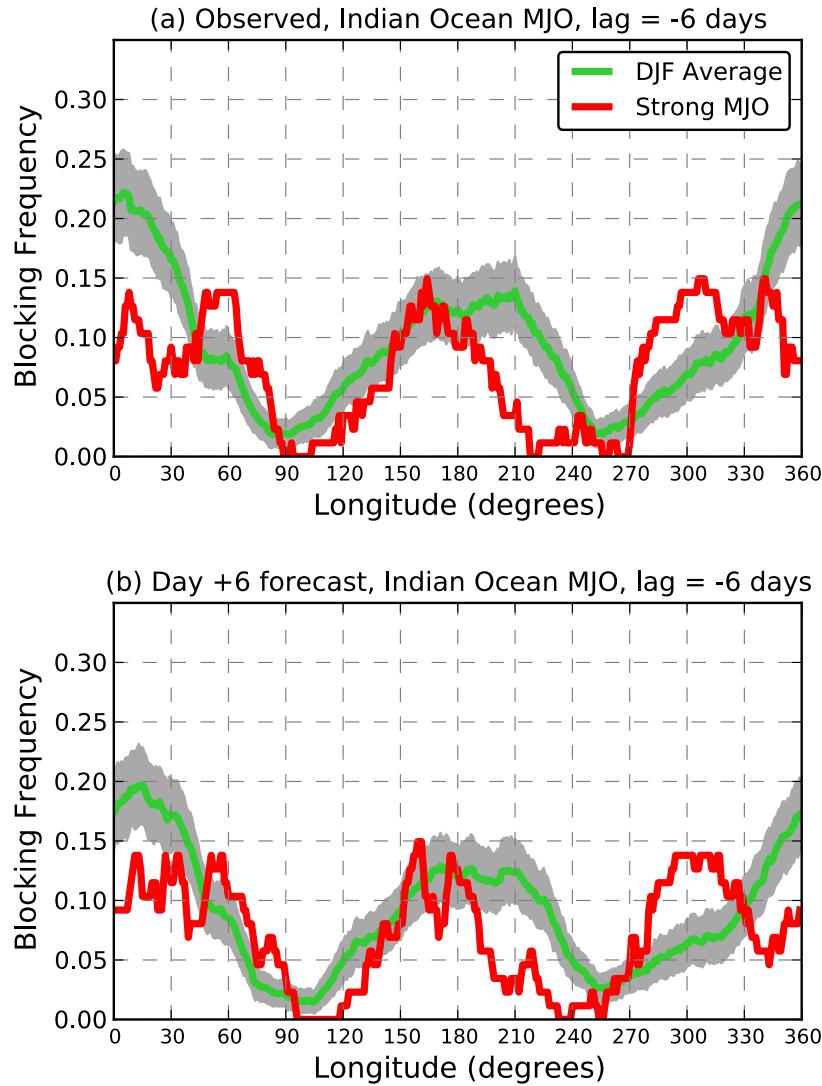
677 September 2005 from the (a) global GFS ensemble reforecast and (b) regional ARW
678 ensemble forecast. The individual ensemble member tracks are shown in gray
679 (control run in green) with red dots marking every 24 hours. The observed track is
680 shown in black with black dots marking every day at 0000 UTC. The inset in (a)

681 shows the intensity forecast for Rita from the global GFS ensemble (gray) and ARW
682 (red). The observed intensity is shown by the blue dashed contour. The black line
683 represents the ensemble mean and the shading encompasses intensity values within
684 the 5% and 95% percentiles. The inset in (b) shows the 48-h forecast composite
685 reflectivity (shaded according to the color bar in dBZ) from the 4-km domain of the
686 control member of the ARW ensemble.

687

688

689



690

691

692 **Figure 7:** (a) Observed, and (b) +6 day forecast blocking frequency as a function of
 693 latitude for December-January-February 1985-2010 (green lines) and for the subset
 694 of cases with an Indian Ocean strong MJO as defined in the text. The MJO data was
 695 defined 6 days prior to the analysis or the forecast. Grey area denotes differences
 696 that are between the 5th and 95th percentile confidence intervals as determined
 697 from a block bootstrap algorithm.

698